DOCUMENT RESUME

ED 460 199 UD 034 649

AUTHOR	Treisman, Philip Uri; Surles, Stephanie A.
TITLE	Systemic Reform and Minority Student High Achievement.
INSTITUTION	Institute of Medicine (NAS), Washington, DC.
PUB DATE	2001-00-00
NOTE	25p.; In: The Right Thing To Do, the Smart Thing To Do: Enhancing Diversity in the Health Professions. Summary of the Symposium on Diversity in Health Professions in Honor of Herbert W. Nickens, M.D., Brian D. Smedley, Adrienne Y. Stith, Lois Colburn, and Clyde H. Evans, Editors. National Academy Press.
AVAILABLE FROM	For full text: http://www.nap.edu/catalog/10186.html.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Academic Achievement; *Black Students; *Educational Change; Elementary Secondary Education; *High Achievement; Higher Education; *Hispanic American Students; *Racial Differences; Standardized Tests; State Standards; Student Evaluation
IDENTIFIERS	National Assessment of Educational Progress; Texas Assessment of Academic Skills

ABSTRACT

The under-representation of African American and Hispanic American students among high achievers on standardized tests, honors graduates of most colleges, and practitioners of mathematics and science professions is well-documented. This paper explores the extent to which the current educational reform movement is achieving the goal of substantially improving overall student achievement, especially high achievement, while also reducing achievement gaps among racial and ethnic groups. Neither high-stakes testing nor standards setting, both components of the modern systemic reform movement, is new, nor are the issues of testing and equity. The primary resource for tracking changes in students' academic performance is the National Assessment of Educational Progress (NAEP). Especially important to evaluating systemic reform efforts is a state NAEP testing program begun in 1990 which allows sophisticated and multidimensional comparisons of state academic performance. While NAEP data show that between 1971-88, the black-white achievement gap closed significantly, 1990s data show that the gaps are growing. There is little hard evidence that state systemic reform policies have had their intended effects. Results from the Texas Assessment of Academic Skills indicate a marked decrease in performance gaps of minority and white elementary students in recent years. (Contains 21 references.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook.php?record_id=10187&page=145.html, copyright 2001, 2001 The National Academy of Sciences, all rights reserved.

ED 460 199

Systemic Reform and Minority Student High Achievement

Philip Uri Treisman and Stephanie A. Surles[1]
University of Texas, Austin

INTRODUCTION

The underrepresentation of African-American and Latino[2] students among the ranks of high achievers on standardized tests, among the honors graduates of most American colleges and universities, and among the practitioners of mathematics- and science-based professions, has been exhaustively documented (Borman, Stringfield, & Rachuba, 2000). This underrepresentation is not limited to the poor, nor is it limited to those with little family history of higher education. The problem is endemic and widespread, and close inspection of the data belies simple explanations.

The dimensions of the underrepresentation problem have been well examined in *Reaching the Top*, a 1999 report of the College Board's National Task Force on Minority High Achievement. According to the task force report, student performance differentials by race and ethnicity appear as early as elementary school and persist through college (College Entrance Examination Board, 1999). The report states that in 1995, African Americans, Latinos, and Native Americans accounted for only 13 percent of the bachelor's degrees, 11 percent of the professional degrees, and 6 percent of the doctoral degrees awarded by U.S. colleges and universities. To put these statistics into perspective, these

[1] The authors wish to thank Catherine Clark and Janis Guerrero for their targeted and thoughtful comments on this paper. We also thank Rachel Jenkins for her keen editorial eye and Brenda Nelson for her timely review.

[2] The terminology used to refer to different ethnic and racial groups varies throughout this paper. We have used the language of the studies we are citing to minimize confusion between categories.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY
K. Barranco
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

UD034649

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook/0309076145/html/261.html, copyright 2001, 2001 The National Academy of Sciences, all rights reserved

groups constituted 30 percent of the under-18 population at that time (College Entrance Examination Board, 1999).

The underrepresentation of minority students is especially severe in mathematics- and science-based majors. This underrepresentation is not due to lack of interest on the part of students. In fact, most high school students heading for college select roughly the same intended majors, regardless of their ethnic background. For example, in a national survey, 21.5 percent of black students, 21.5 percent of Hispanic students and 20.9 percent of white students selected science or engineering as their intended majors in 1998. This data includes majors in the natural sciences, engineering, and mathematics and computer sciences (National Science Board, 2000).

The task force also documented very low percentages of high performers among minority students on various tests, including the SAT I and the National Assessment of Educational Progress assessments in reading, mathematics, and science. Critics of standardized testing often attribute these low percentages of minority students among high achievers—and the low average scores of certain minority groups—to cultural and economic biases or other deficiencies of the tests.[3] Although it is well documented that family income levels are positively correlated with student achievement levels, low minority academic performance is not restricted to the children of low-income families. In 1998, the mean SAT I scores of white and Asian students in low-income families, defined as families earning less than $20,000, were *higher* than the mean SAT I scores of African-American students in high-income families, defined as families earning more than $60,000 (Gandara & Maxwell-Jolly, 1999). Thus, variation in family income is only one of several factors that affect test scores, and the contribution of these factors to minority test scores appears to be less than is often presumed.

The predictive value of test scores is another issue that is more complex than generally thought. One would expect that given the impoverished primary and secondary schools many minority students attend, these students would so flourish academically once they reach the fertile soil of higher education that their performance in college would be better than their SAT I scores would predict. Instead, the SAT I scores of minority students overpredict performance at traditionally white colleges and universities (Bowen & Bok, 1998; College Entrance Examination Board, 1999). In other words, within the same range of SAT I scores, the average college grades of minority students are lower than the average college grades of white students.

The overall evidence on the correlates and predictive value of test scores suggests a pernicious problem—namely, that the forces impeding the academic achievement of minority students persist and take new forms at each level of

[3] See, for example, the work of FairTest: The National Center for Fair & Open Testing, at www.fairtest.org.

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook.php?isbn=0309076145&chap=2-2.html, copyright 2001, 2001 The National Academy of Sciences, all rights reserved

schooling. These counterforces to minority academic advancement remain poorly understood and must become a major focus of American educational research.

Because this country's population is undergoing a demographic shift, and because this shift is even more dramatic in the school-aged population, the difference in performance among students of different ethnic and racial groups is growing in social consequence. If the inequity in educational opportunity that exists today is not addressed, particularly the opportunity that nurtures high levels of achievement, then this inequity will become an increasingly powerful source of division and inequality.

Furthermore, given recent limitations driven by state policy and court rulings on the use of race and ethnicity as factors in undergraduate and graduate school admissions, without a dramatic narrowing of the achievement gap among racial and ethnic groups, the nation's professional and political leadership could grow less and less demographically representative of the population it serves. In both California and Texas, for example, the law schools of the states' flagship public research universities have graduated individuals who constitute a large percentage of each state's minority legislators.[4] The question then arises: will professional and political leadership in a pluralistic democracy retain its authority if it does not truly reflect the diversity of the population it serves?

MODERN SYSTEMIC REFORM AND HIGH-STAKES TESTING

Our purpose in this paper is to explore the extent to which the current American educational reform movement is achieving the important goal of substantially improving overall student achievement, and especially high achievement, while at the same time reducing achievement gaps among racial and ethnic groups. The fundamental presumption of modern educational reform is that all children should have equal access to high-quality contemporary curricula, and that all students (or in practical terms, nearly all) should be expected to master the content of these challenging curricula and related complex problem-solving skills. Modern reform ideas, as articulated by scholars and policy analysts representing a broad cross-section of political thought, justify this presumption both in economic terms, as necessary for our international economic competitiveness, and in terms of basic justice, as a natural extension of the American equity creed.

Modern systemic reform frameworks typically consist of four interconnected mechanisms designed to increase educational system coherence in the

[4] California is effected by Proposition 209, which went into effect on August 28, 1997. It is now Section 31 of Article I of the California State Constitution. Texas, Louisiana, and Mississippi are affected by the decision of the U.S. Court of Appeals for the Fifth Circuit in *Hopwood* v. *Texas*, 78 F.3d 932 (5th Cir. 1996).

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook/0309087651/45/html/263.html. copyright 2001, 2001 The National Academy of Sciences. all rights reserved.

belief that such coherence will lead to greater student achievement. The first mechanism consists of three subsystems: a curriculum guidance system driven by state curriculum documents that define what children should know and be able to do, an assessment system that measures student progress in learning content delineated in the curriculum documents, and an accountability system in which schools are rewarded or sanctioned based on their students' performance. The curriculum documents that drive the system should ideally be developed through a participatory process that balances the judgments of scholars and professional educators with those of the business leaders and members of the public at large. This tripartite first mechanism is most commonly thought to be synonymous with systemic reform, but most policy analysts believe that it alone is not enough to make lasting or significant change. The following three additional mechanisms are also essential to the success of systemic reform efforts.

The second mechanism of modern systemic reform is the alignment of state policies concerning textbook adoption; administrator certification; and teacher preparation, licensure, and continuing education, as well as the panoply of related issues that shape who teaches and what gets taught. In systemic reform, each of these domains is organized or reorganized to align with the state's adopted curriculum. This alignment ensures, at least in principle, that all major efforts and expenditures are focused on the same end.

The third mechanism is a restructured educational governance system that in exchange for greater state-level authority over curriculum, grants local schools and school districts greater autonomy in local school management—that is, more flexibility in rules for hiring, in making expenditures, and so on. In essence, then, modern systemic reform gives up certain forms of state control in exchange for the state's greater authority over curriculum. This exchange of power is necessary in a reform agenda focused on the substance of what children learn. In the words of Marshall Smith, one of the architects of standards-based reform, "this model of content-driven systemic reform would . . . marry the vision and guidance provided by coherent, integrated, centralized education policies common in many nations with the high degree of local responsibility and control demanded by U.S. tradition" (O'Day & Smith, 1993).

The fourth mechanism, which has received relatively little attention in current reform efforts, is a strategy for ensuring the availability of resources necessary to bring about desired changes in learning. Smith argued that without adequate resources available to schools that serve poor and minority students, higher standards would lead to an increase in the achievement gap and greater levels of failure for minority and poor children. Adequate resources, including qualified teachers, adequate facilities, and proper instructional materials, constitute what is often called "opportunity to learn" (Rothstein, 2001). Certainly, the determination of what constitutes an adequate level of resources is at the center of one of the liveliest debates in American education policy and politics (Guthrie & Rothstein, 2001).

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook/0309070147/html/264.html, copyright 2001, 2001 The National Academy of Sciences, all rights reserved

The humbling challenge facing proponents of systemic reform, or of any other broad reform strategy, is the essential localism of American education. Although the modern American education system and its ideas of positive governmental responsibility for education is largely a post-Civil War construction, the tradition of local governance is deeply grounded in American history and culture. Its roots can be traced to the sentiments of the country's founders, who came to America seeking relief from religious persecution by European governments (Tyack, 1974; Ravitch, 1995). This venerable tradition of local control now finds support both among conservatives, who oppose on principle expanded roles of the federal government, and among liberals, who fear that a national system based on high standards might reinforce existing social stratification and devalue minority cultures. Conservative and liberal critics alike fear that a national curriculum would, of necessity, be a product of compromises that could lead to a sterile and narrow curriculum and to impoverished, test-driven pedagogy. The deeply held belief that those closest to children should determine the content of their education has produced a national educational enterprise with an almost immunological ability to resist centralized reform initiatives.

American education today consists of 50 state systems, each composed of multiple school districts[5] governed by locally elected school board members who do not report in any substantive way to a state board of education. Further, these school districts are managed by superintendents who do not report formally to the state superintendent. Thus, movements to modernize or reform schools must gain the support of the citizenry if they are to affect local education practice. In some sense, then, the effectiveness of national reform strategies can be assessed by the robustness of their strategies for inculcating coherence in a system designed to resist such coherence.

HISTORICAL ANTECEDENTS OF SYSTEMIC REFORM STRATEGIES

Neither high-stakes testing nor standards setting—components of the modern systemic reform movement—is new. With roots that go back at least two millennia to the civil service examination system of the Han Dynasty (206 B.C.–220 A.D.), high-stakes testing has a long and venerable history as a driver of educational systems and as a controller of who, by virtue of their education, can have access to power. The issues of testing and equity are not new, either. By the 13th century, China had created a broad-based system of public schools explicitly designed to prepare the talented but indigent for the high-stakes civil service examinations.

During the Ming Dynasty (1368 A.D.–1644 A.D.), China's civil service examination system evolved in ways that today might be called standards-based.

[5] With the exception of Hawaii, which has only one statewide school district.

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook/0309067014S/html/265.html, copyright 2001, 2001 The National Academy of Sciences, all rights reserved

Examination content was exclusively drawn from the Nine Classics of Confucianism and an established form emerged for the questions and the required answers.[6] The Chinese civil service examination system, which was abolished at the beginning of the 20th century, strongly influenced the British civil service examination system, as well as those of other former European colonial empires.

The notion of standards setting also has very early roots in American education, as evidenced by Webster's "Blue-backed" speller (1782), which taught American children how to read, spell, and pronounce words (Ravitch, 1995). Testing also has a long history in America, with roots going back to the 1890s. That decade marked the beginning of standardized testing to measure learning outcomes by school and the development of intelligence tests to measure students' mental ability. By the 1920s, both forms of testing were powerful forces in big city school systems and were the basis for ability differentiated learning groups, or curricular "tracking"—a practice that remains commonplace in U.S. education and that remains a major concern for advocates of educational equity.

In 1900, the College Entrance Examination Board (CEEB) created the first coherent American system of high-stakes testing designed to directly influence school curriculum, albeit in a small and specialized collection of high schools (Valentine, 1987). At that time, independent colleges and universities paid little attention to students' prior academic achievement in their admissions decisions. Instead, they focused on ability to pay, religion, and social status. In many cases, wealthy families enrolled their children in college at birth. Before 1900, each selective Ivy League college used its own admissions test. Tests were sufficiently different from one another to create formidable difficulties for any high school seeking to prepare students to gain entry to any one of a broad array of colleges. The premise underlying the CEEB system was that by making the standards for college admission clear, and by using these standards to create a common college entrance examination, high schools with good teachers could reliably prepare diligent students for admission to selective Ivy League colleges. This same premise—that good teaching and student diligence rather than extraordinary talent or privilege are the essential prerequisites for academic success—is an axiom of modern educational policy. Under the CEEB system, however, children from wealthy families had an additional advantage, as paid exam-preparation tutors could help ready students for college entrance examinations.

The CEEB system contained many features of interest to today's reformers. The College Board developed a novel and vibrant process for creating and continually revising the standards on which its examinations were based. These standards were delineated in annual publications called the "Definitions of the Requirements," and represent the result of often intense debates about "what students should know and be able to do." The College Board also created a reli-

[6] Examination papers had eight main headings, and each answer was to be structured in a highly prescribed manner and limited to fewer than 700 characters.

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook.php?isbn=0309072654&chap=266.html, copyright 2001, 2001 The National Academy of Sciences, all rights reserved

able process for turning standards into examinations. Of special power were the readings, a practice in which college faculty and, especially in the later years of the system, high school teachers, convened to grade the examinations. This practice became an exemplar of effective professional development for teachers.

From its earliest days, critics and even members of the College Board were concerned that the college entrance examination system was exerting too large an effect on the high school curriculum. Further, the rapid advances of modern psychology and the social sciences in the first decades of the 1900s, and the broad use of intelligence testing during World War I, led the College Board to explore the use of intelligence testing as a tool for college admissions. The result was the Scholastic Aptitude Test, a precursor to today's SAT, which was first administered in 1926 to 8,040 students (Lemann, 1999). In the 1930s, the rapid advance of psychology, which then held that intelligence was largely innate and easily measured, and the fears of class warfare during the Great Depression led to the rapidly increasing use of the SAT. Elite colleges were seeking to broaden the socioeconomic diversity of their student bodies and needed a mechanism to identify unpolished talent. Support for the SAT grew among college officials as a replacement to the CEEB examination system.

On December 7, 1941, immediately after learning of the attack on Pearl Harbor, the admissions directors of Harvard, Yale, and Princeton, who were meeting about College Board issues, decided that candidates for admission to their campuses would not take the June CEEB. Instead, they would take the April SAT, so that students could start taking courses in the summer rather than having to wait for the fall (Valentine, 1987). Other colleges quickly followed suit, and the College Board cancelled the June CEEB examinations, effectively ending the program. The SAT remained, and until the creation of the American College Testing Program (ACT) in 1959, was the only national examination used for college admissions.

Historians view the SAT as a primary instrument in bringing about the diversification of American higher education, at least for Catholics and Jews. Donald Stewart, a former president of the College Board, has long argued that under affirmative action the SAT was also instrumental in diversifying the ethnic and racial makeup of American higher education. Certainly, for at least three decades beginning in 1960, the SAT and the ACT were used by selective public and private colleges and universities to identify minority students who they believed, by virtue of their test scores, had the potential to succeed. In places where affirmative action policies in college admissions have been ended, it can conversely be argued that the SAT is now a hurdle to minority access to higher education because, as described above, relatively few minority students receive high scores on standardized tests. More broadly, the question of which approach to testing best advances equity—one, like the SAT, that purports to identify un-

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook.php?isbn=0309075165&lpage=267.html. copyright 2001, 2001 The National Academy of Sciences. all rights reserved

varnished talent, or one, like the CEEB, that purports to test mastery of a defined curriculum—remains contentious and unresolved (Atkinson, 2001).[7]

The compensatory education movement of the mid-1960s to the early 1980s provides a powerful example of using testing to shape educational practice. This movement was focused explicitly on raising the academic performance of poor children and on reducing the gap in test scores between minority and non-minority students. The instructional focus of the movement was on reducing illiteracy and innumeracy and, more generally, on ensuring that all students learned basic skills as measured by national norm-referenced tests. Federal legislation, particularly the Elementary and Secondary Education Act (ESEA), provided great impetus to the testing movement by requiring regular testing in exchange for federal funding for disadvantaged students. Tests, textbooks, and teacher professional development in high-poverty communities were mutually reinforcing. Many states adopted minimum competency testing and, going beyond federal requirements, included such testing as part of their high school graduation requirements (Texas Education Agency, 1996). The use of testing in Title I of ESEA provided a model for the use of testing in state education reforms of the 1990s (Ravitch, 1995).

The compensatory education movement did, in fact, help increase the test scores of minority and economically disadvantaged children. Because the test scores of relatively advantaged children were comparatively stable during this period, the disparities in scores were dramatically reduced (O'Day & Smith, 1993). NAEP data show that from 1971 to 1988, achievement gaps between African-American and white students were reduced by 30% to 60%, depending on the subject tested. There is some dispute, however, about how much of this performance narrowing was the result of instruction and how much was the result of significant reductions in the percentage of children living in poverty—an issue of fundamental concern to both researchers and policy analysts (O'Day & Smith, 1993).

A LOOK AT THE DATA

National Assessment of Educational Progress

The primary resource for tracking changes in students' academic performance is the National Assessment of Educational Progress, also known as NAEP,

[7] Concerns about the SAT and similar exams have now so permeated American society that the University of California system, among others, is wrestling with the question of whether the tests have become too dominant a force. The concern is that the tests are limiting the curriculum to such an extent that they are harming the education of children. The president of the University of California system has recommended that the university system no longer require the SAT I. A complete description of University of California president Richard C. Atkinson's views on the SAT I are in the 2001 Robert H. Atwell Distinguished Lecture which he delivered at the 83rd Annual Meeting of the American Council on Education.

BEST COPY AVAILABLE

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook/0309076145/html/268.html, copyright 2001, 2001 The National Academy of Sciences, all rights reserved

or the "Nation's Report Card." Established in 1969, NAEP now includes a variety of assessment programs designed to measure changes over time in students' competence in basic skills as well as in the content defined by important national standards documents (such as that published by the National Council of Teachers of Mathematics[8]). Especially important to evaluating systemic reform efforts is a state NAEP testing program begun in 1990, which allows sophisticated and multidimensional comparisons of state academic performance. Data from this NAEP assessment can be used to compare student performance from state to state at grades 4, 8, and 12. This NAEP data set can also be used to estimate the learning gains of cohorts of students as they move, for example, from the fourth to the eighth grade.

To illustrate how NAEP supports complex analyses of state performance, we examine data from two states with roughly similar student populations, California and Texas. In Figure 1, we report the states' fourth and eighth grade average mathematics scale scores in 1992 and 1996, the most recent years for which NAEP data are available. In a cohort analysis—comparing the scores of 1996 eighth graders with those of 1992 fourth graders—California appears to have done a slightly better job than Texas in supporting ongoing student mathematics learning. If, however, one compares state scores at fixed grade levels—i.e., 1992 fourth graders with 1996 fourth graders and 1992 eighth graders with 1996 eighth graders—it appears that Texas has done a better job in supporting student learning. Note that the absolute differences in the performance of Texas and California students are strikingly large—20 points in 1996 at the fourth-grade level and 7 points at the eighth-grade level. A difference of about 11 points corresponds to one year of learning (Loveless & Diperna, 2000).

NAEP allows the disaggregation of a state's performance data by major ethnic or racial subgroup, a crucial fact for its contributions to analysis of performance. As one example of the power of this capacity, we report in Figure 2 the average scale scores of participating states[9] on the 1998 eighth-grade NAEP writing assessment. African-American students in Texas score as well as or bet-

California	4th Grade	8th Grade
1992	208	261
1996	209	263

Texas	4th Grade	8th Grade
1992	218	264
1996	229	270

FIGURE 1 1992 and 1996 NAEP Mathematics Scores for California and Texas
SOURCE: NAEP 1996 Mathematics Report Card for the Nation and the States.

[8] These standards are available online at http://standards.nctm.org/.

[9] Currently, state NAEP assessments are voluntary, but the great majority of states choose to participate.

BEST COPY AVAILABLE

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions in
http://www.nap.edu/openbook/0309067111/html/269.html. copyright 2001, 2001 The National Academy of Sciences. all rights reserved

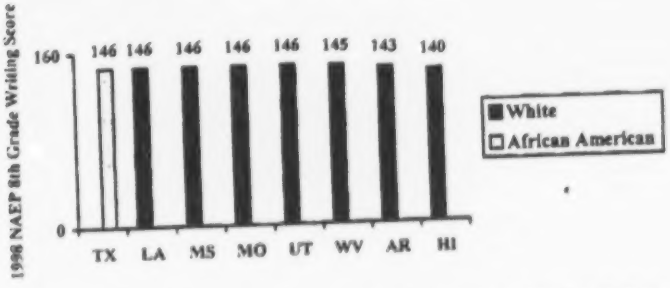

FIGURE 2 1998 NAEP 8th Grade Writing for Texas African-American Students and White Students in Seven Other States.
SOURCE: National Center for Educational Statistics and The Education Trust.

ter than white students in seven other states. (These various analyses of the data dispel the myth that the demography of a state determines the outcomes of its educational system. After all, it is the adults in the system, not the children, who determine its outcomes.)

As described above, NAEP data show that nationally, between 1971 and 1988, the gap in scores between African-American and white students closed significantly. Figure 3 shows the closing of the gap in NAEP reading scores. There was also a substantial closing of the gap between the scores of middle class children and those of poor children. There is no evidence, however, that these gaps have closed further in the 1990s. Rather, the unhappy finding is that nationally the gaps, especially in reading, are growing (Barton, 2001). The positive news is that there have been gains in average scores for all the major subgroups tested; even so, nationwide these gains can be fairly described as being small to modest (Barton, 2001). The data clearly indicate that younger students have made significantly greater progress than have older students (Loveless & Diperna, 2000).[10] Indeed, the increases in reading and mathematics learning gains from grades 8 to 12 are about one-third the gains from grades 4 to 8.

[10] One of the few findings about which there is broad agreement is that learning gains decrease as students move through the American education system. This deceleration of learning gains once students reach high school has profound consequences for equity in and access to mathematics- and science-based careers.

BEST COPY AVAILABLE

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook/0309076145/html/270.html, copyright 2001, 2001 The National Academy of Sciences, all rights reserved

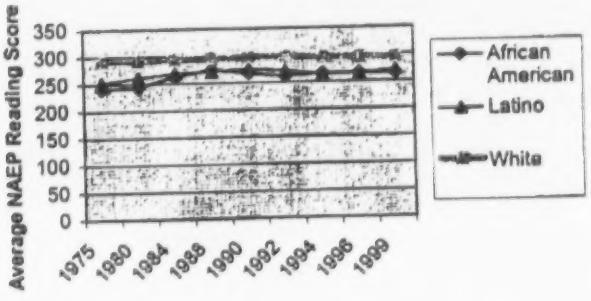

FIGURE 3 Gap in NAEP Reading Scores for 12th Graders.
SOURCE: U.S. Department of Education, National Center for Educational Statistics and The Education Trust.

Of course, analysis of NAEP data alone cannot tell us how much of the observed gain in student performance in a state is the result of that state's systemic reform strategy. In the 1980s and 1990s, states implemented a broad array of policies addressing what were believed to be important opportunities for improving learning. These included state initiatives to lower class size (especially in the early grades), to raise teacher and administrator salaries, to strengthen teacher licensing requirements, to create mentoring programs for new teachers, and so on.

In their analysis of NAEP mathematics gains, David Grissmer and Ann Flanagan controlled for the most common of these strategies and still found that modest progress—at least in mathematics—can be attributed to state systemic reform policies (Grissmer & Flanagan, 1998). Grissmer and Flanagan found that Texas and North Carolina, early adopters of standards-based reform, show the largest gains.

In a separate study, Grissmer and Flanagan used census figures to adjust NAEP data that rely on student reports of key variables—such as family income—about which children may be poor judges (Grissmer & Flanagan, 2001). Grissmer and Flanagan then disaggregated adjusted state NAEP data by geographic locality—urban, suburban, and rural—to see the breadth of effect that state policy has on student learning gains. The researchers hypothesize that "achievement gains within states should occur in all localities within the state . . . if statewide reforms are effective" (Grissmer & Flanagan, 2001). Using this test and their adjusted data, they find that only three states meet this criterion of broad geographic effect—Michigan, North Carolina, and Texas.

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook/0309076145/html/271.html, copyright 2001, 2001 The National Academy of Sciences, all rights reserved

These three states agreed to participate along with ten others in TIMSS-R,[11] a benchmarking project conducted as part of the 1999 Third International Mathematics and Science Study (Jerald, 2001). In Table 1, we show data on the relative performance of these states, as well as explicit benchmarking data relevant to high performance. The data show that Michigan and Texas have the highest average TIMSS scores; Texas also has the highest percentages of students scoring at the top reported levels of performance. Note that, as can also be seen in Table 1, Texas had substantially higher proportions than did other states of minority and low-income test takers.

A natural extension of the disaggregation idea is to examine student learning gains not only by geographic locality, but also by the race or ethnicity of students and by subject matter (for example, gains in reading versus mathematics). If policy was the primary driver of improvement, then Grissmer and Flanagan's argument would suggest that performance would be somewhat uniform across these three policy relevant variables. Subject matter is important as a check on the effects of state policy because mathematics and reading are typically treated with equal or nearly equal weight in state policy frameworks and testing programs. Ethnicity and locality are both important because they illuminate a canonical legislative process—that is, any legislature's need to balance the competing interests of rural, urban, and suburban constituencies, as well as ethnic and racial constituencies. Finally, disaggregation by ethnicity is especially important because the *raison d'etre* of systemic reform is to achieve equity in the educational system.

A look at state NAEP data in the 1990s shows that no state at any grade level has shown statistically significant gains across all academic subjects, ethnic and racial groups, and localities. Grissmer and Flanagan were surprised by the apparent near statistical independence of these variables. Between 1992 and 1998, 13 states made statistically significant gains on the fourth-grade NAEP reading assessment; between 1992 and 1996,[12] 16 states made statistically significant gains on the fourth-grade NAEP mathematics assessment (Grissmer & Flanagan, 2001). Only six states showed statistically significant gains in both subjects.

If we examine NAEP reading data for differences by locality—that is, rural, urban, or suburban—only Connecticut made statistically significant gains in two of the three localities. No other state made gains in more than one locality. Of

[11] This is the Third International Mathematics and Science Study-Repeat.

[12] Each time NAEP is administered it does not test students on every academic subject; thus, the variation in dates for reading and mathematics.

BEST COPY AVAILABLE

TABLE 1 1999 TIMSS-R 8th Grade Mathematics Scores

Average Math Score		Students Scoring in Top 10% Internationally		Students Scoring in Top 25% Internationally		Minority Student Test Takers		Low-Income Student Test Takers	
Michigan	***517***	***Texas***	***13%***	***Texas***	***37%***	***Texas***	***53%***	***Texas***	***48%***
Texas	***516***	Connecticut	11%	***Michigan***	***33%***	Maryland	45%	South Carolina	45%
Indiana	515	Illinois	10%	Oregon	32%	North Carolina	38%	North Carolina	44%
Oregon	514	Massachusetts	10%	Connecticut	31%	South Carolina	37%	Idaho	37%
Massachusetts	513	***Michigan***	***10%***	Massachusetts	31%	Illinois	35%	Missouri	34%
Connecticut	512	Oregon	10%	Indiana	30%	Connecticut	26%	Oregon	33%
Illinois	509	South Carolina	10%	South Carolina	30%	Massachusetts	26%	Illinois	31%
Pennsylvania	507	Indiana	9%	Illinois	29%	Missouri	22%	Pennsylvania	30%
South Carolina	502	Pennsylvania	9%	Pennsylvania	28%	Pennsylvania	22%	Massachusetts	28%
North Carolina	495	Maryland	8%	Maryland	27%	Oregon	20%	Maryland	28%
Idaho	495	North Carolina	7%	North Carolina	25%	***Michigan***	***18%***	Indiana	25%
Maryland	495	Idaho	5%	Idaho	24%	Idaho	17%	Connecticut	20%
Missouri	490	Missouri	4%	Missouri	20%	Indiana	17%	***Michigan***	***17%***

SOURCE: International Study Center and The Education Trust

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions in
http://www.nap.edu/openbook.php?isbn=0309074355&chap=272.html. Copyright 2001, The National Academy of Sciences. All rights reserved.

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook.php?isbn=0309070147&page=273.html, copyright 2001, 2001 The National Academy of Sciences, all rights reserved

Grissmer and Flanagan's top tier of states, Texas showed statistically significant gains in reading only for rural students; no locality in North Carolina or Michigan showed statistically significant gains.

NAEP also reports proportions of students in each state who score at the four NAEP-defined levels: below basic, basic, proficient, and advanced. The data show that in no state in either 1992 or 1996 did more than 3% of students score at the advanced level. The data for minority students are even worse—almost no minority students in any state score at the advanced level. Journalist and researcher Richard Rothstein has pointed out that NAEP levels are set arbitrarily, so that low percentages of students (regardless of ethnicity) scoring at the advanced level may be an indication of unrealistic level-setting (Rothstein, 1999). Regardless, the fact that low numbers of minority students are achieving at very high levels is a special problem for the development of minority leaders.

In overview, then, we find little hard evidence that state systemic reform policies have had their intended effects. Nonetheless, there is some supporting evidence for attributing gains in mathematics learning in Texas, Michigan, and North Carolina to their state policy frameworks. There is also evidence that gains in minority achievement in Texas can be at least partially attributed to the use of disaggregated data in its accountability system (Treisman & Fuller, 2001). To examine educational performance in Texas, we turn to data from the Texas Assessment of Academic Skills (TAAS), the principal state testing instrument.[13]

Texas Assessment of Academic Skills

Texas has a comparatively long history with systemic reform and has made an unusually large investment in collecting data on school and school district performance, including student test performance (Texas Education Agency, 1996). Moreover, because of a recent high-profile legal challenge to the state's testing system,[14] and a Texan's prominence in the 2000 presidential election race, these Texas data have received careful scrutiny.

In Figures 5 and 6, we examine differences in TAAS passing rates of white and African-American students and of white and Hispanic students on both the mathematics and reading portions of the TAAS. As the data clearly show, there has been a marked decrease in the performance gaps in the elementary school years. Note that state reading results are far less impressive, mirroring the na-

[13] The TAAS measures the statewide curriculum in reading and mathematics at grades 3–8 and 10; in writing at grades 4, 8, and 10; and in science and social studies at grade 8. Satisfactory performance on the TAAS 10th grade tests is a prerequisite to a high school diploma. TAAS scores are also used to rate individual schools and school districts.

[14] See *GI Forum et al.* v. *Texas Education Agency et al.* at 87 F. Supp. 2d 667. The issue in this case was whether the use of Texas' high-stakes exam as a requirement for high school graduation unfairly discriminated against Texas minority students or violated their right to due process. The court found that it did neither. Uri Treisman served as an expert witness for the Texas Education Agency in this case.

BEST COPY AVAILABLE

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions in
http://www.nap.edu/openbook.php?isbn=0309074145&page=274.html, copyright 2001, 2001 The National Academy of Sciences. all rights reserved.

tional pattern, despite considerable political attention and substantial state funds dedicated to improving reading performance. Similar data exist for the middle school years. It is important to note that these analyses are possible because Texas has required TAAS data to be disaggregated by ethnic and racial group.

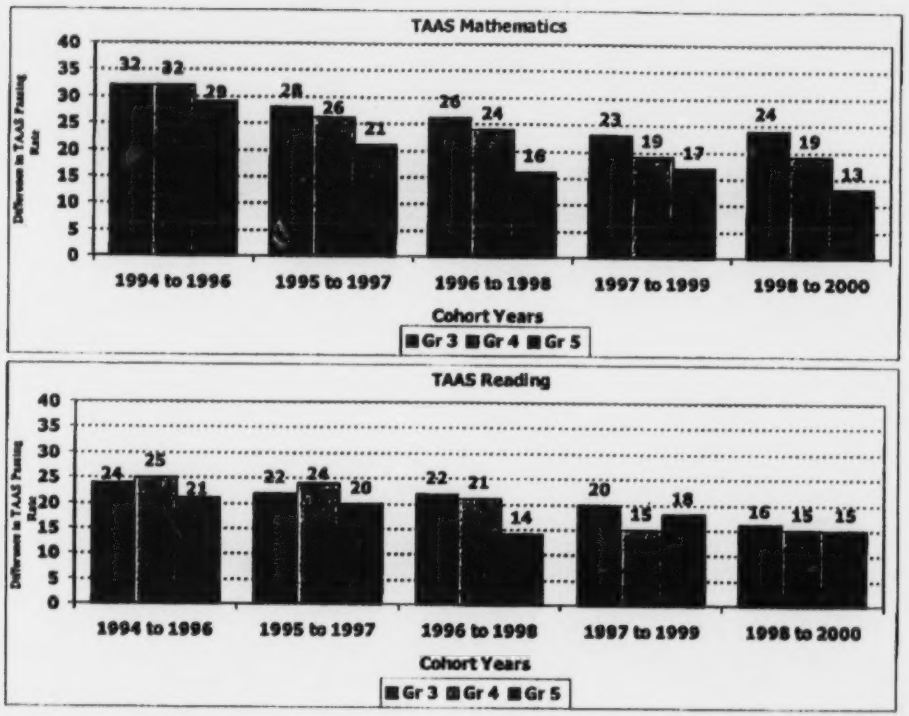

FIGURE 5 Cohort Analysis of Performance Gaps on TAAS Between Texas White and African-American Elementary Students.
SOURCE: Texas Education Agency and Edward J. Fuller

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook/0309067715/html/275.html, copyright 2001, 2001 The National Academy of Sciences all rights reserved

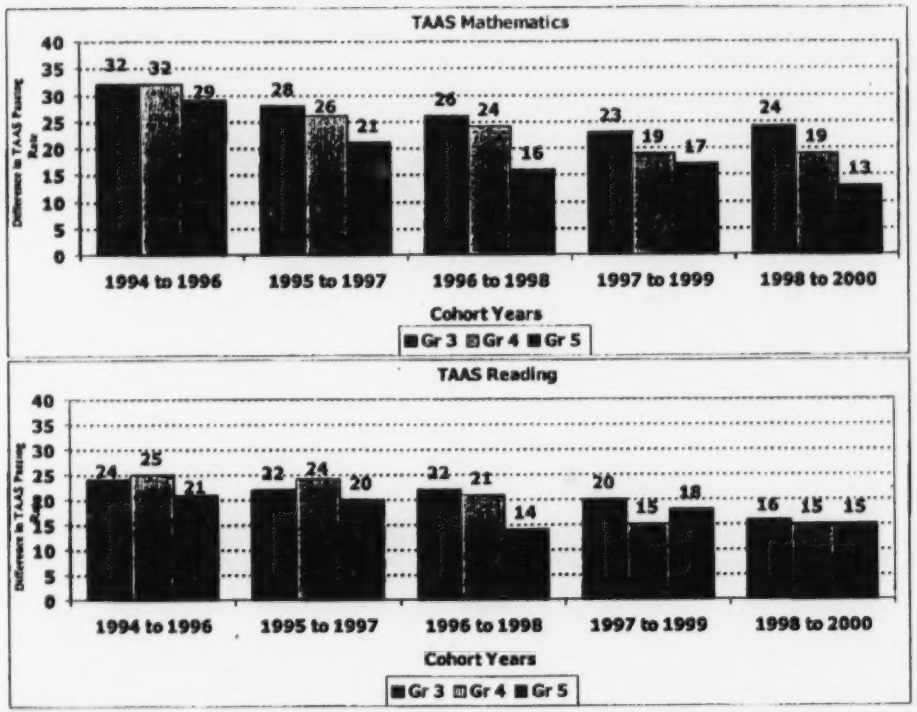

FIGURE 6 Cohort Analysis of Performance Gaps on TAAS Between Texas White and Hispanic Elementary Students.
SOURCE: Texas Education Agency and Edward J. Fuller.

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook.php?isbn=0309071437&page=270, copyright 2001, 2001 The National Academy of Sciences all rights reserved

BEST COPY AVAILABLE

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook.php?isbn=0309076145&html=277.html, copyright 2001, 2001 The National Academy of Sciences, all rights reserved

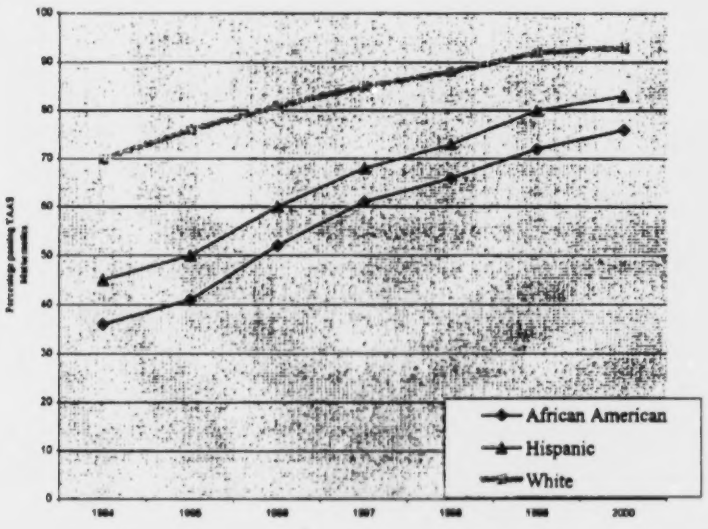

FIGURE 7 Students passing TAAS Mathematics Grades 3–8 and 10.
SOURCE: Texas Education Agency.

We see in Figure 7 that ceiling effects on the testing instrument may be responsible for some of this decrease in the performance gaps.

A close look at Texas education between 1980 and 2000 suggests that for systemic reform to have its intended effect, schools and school systems need to have certain capacities and supports (Treisman & Fuller, 2001). In Texas, a series of successful court challenges to the state's system of financing education addressed massive inequities in local capacity to offer the kind of education envisioned in the state's standards.[15] Rural and urban schools, as well as schools with high proportions of poor children, received large increases in state aid that allowed them to create "opportunities to learn." Further, a broad-based coalition of business leaders heavily engaged in strengthening Texas education through a standards-based approach counterbalanced the natural legislative tendency to

[15] In the case of *Edgewood Independent School District et al.* v. *Kirby et al.* (777 S.W.2d 391), the Texas Supreme Court ruled that the Texas school financing system was unconstitutional under the Texas constitution, which makes education a fundamental right for each child. A series of cases were generated around this issue.

BEST COPY AVAILABLE

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook.php?isbn=0309076145&html=278.html, copyright 2001, 2001 The National Academy of Sciences, all rights reserved

enact short-term solutions to long-term problems (Treisman & Fuller, 2001).[16] In addition, a broad-based mathematics education leadership, by taking an ecumenical approach to curriculum and pedagogy, prevented the dissipation of teacher energy associated with the politicization of the curriculum (Treisman & Fuller, 2001). It would be hard to argue that policies concerning standards, testing, and school accountability alone are responsible for observed improvements in Texas's mathematics education.

LOOKING TO THE FUTURE

Perhaps the systemic reform movement, with its powerful armamentarium of explicit performance standards, tests tied by law to these standards, and systems of rewards and sanctions for schools and districts, can, under the best circumstances, significantly reduce the low achievement that has been endemic in schools serving largely minority and poor populations. But, if high achievement is to become a common outcome of our educational systems, we must develop comprehensive policies that make such achievement a priority. These policies must be supported by substantial federal, state, and local investments so that the accident of where children attend school does not determine the academic opportunities and challenges that they can pursue. Developing such policies, and the wherewithal to fully implement them, will require the support of leaders in all of those sectors that, in the end, determine the priorities of government spending and of local action.

To sustain a system that produces high achievement, we will need to develop better mechanisms for measuring and reporting such achievement and for using that information to strengthen our schools. We must pay special attention to support systems for teachers and administrators, so that the nurturance of their students' intellectual and artistic passions becomes central to the curriculum. We must invest in building the extracurricular institutions that have produced so many of today's senior professionals (College Entrance Examination Board, 1999), and we must charge these institutions with building a generation of professionals that reflects our increasingly diverse population.

American educational policy need not choose between the elimination of low achievement and the promotion of high achievement. We have in our history learned how to achieve both of these goals. The imperative now is to attain them simultaneously.

---

[16] An important state policy that is beyond the scope of this paper is Texas' strategy of incremental improvement. Originally, the standards for the educational system were set at a relatively low level so that most schools could meet them with only modest increases in resources. The standards were then "ratcheted up" over time.

BEST COPY AVAILABLE

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook/0309076145/html/279.html, copyright 2001, 2001 The National Academy of Sciences, all rights reserved

REFERENCES

Atkinson, R. (2001 February 18). Standardized Tests and Access to American Universities. Robert H. Atwell Distinguished Lecture presented at the 83rd Annual Meeting of the American Council on Education. Washington, DC: American Council on Education. [Online]. Available: http://www.ucop.edu/ucophome/pres/comments/satspch.html.

Barton, P. (2001). Raising achievement and reducing gaps: Reporting progress toward goals for academic achievement. Commissioned paper. Washington, DC: National Education Goals Panel.

Borman, G., Stringfield, S., & Rachuba, L. (2000). Advancing minority high achievement: National trends and promising programs and practices. New York: College Entrance Examination Board.

Bowen, W., & Bok, D. (1998). The shape of the river: Long-term consequences of considering race in college and university admissions. Princeton, NJ: Princeton University Press.

College Entrance Examination Board. (1999). Reaching the top: A report of the National Task Force on Minority High Achievement. New York: College Entrance Examination Board.

Gandara, P., & Maxwell-Jolly, J. (1999). Priming the pump: Strategies for increasing the achievement of underrepresented minority undergraduates. New York: College Entrance Examination Board.

Grissmer, D., & Flanagan, A. (1998). Exploring rapid score gains in Texas and North Carolina. Commissioned paper. Washington, DC: National Education Goals Panel.

Grissmer, D., & Flanagan, A. (2001). Searching for indirect evidence for the effects of statewide reforms. Brookings Papers on Education Policy 2001. D. Ravitch (Ed.). Washington, DC: Brookings Institution Press, 2001.

Guthrie, J., & Rothstein, R. (2001). A new millennium and a likely new era of school finance. S. Chaikind & W.J. Fowler, Jr. (Eds.). Education Finance in the New Millennium. Larchmont, NY: Eye on Education.

Jerald, C. (2001). Real results, remaining challenges: The story of Texas education reform. Commissioned paper. Washington, DC: The Business Roundtable.

Lemann, N. (1999). The big test: The secret history of the American meritocracy. New York: Farrar, Straus and Giroux.

Loveless, T., & Diperna, P. (2000). How well are American students learning? Focus on math achievement. The Brown Center Report on American Education, Volume I, Number 1. Washington, DC: The Brookings Institution.

National Science Board. (2000). Science and engineering indicators 2000. Washington, DC: National Science Foundation.

O'Day, J., & Smith, M.. (1993). Systemic reform and educational opportunity. In Designing coherent education policy: Improving the system, S.H. Fuhrman, (Ed.) San Francisco: Jossey-Bass Publishers.

Ravitch, D. (1995). National standards in American education: A citizen's guide. Washington, DC: Brookings Institution Press.

Rothstein, R. (April 12, 1999). NAEP gets failing grade on grading. The American Prospect. [Online]. Available: http://www.epn.org/idea/columns/ro981021.html.

Rothstein, R. (2001). Comment by Richard Rothstein. In Brookings papers on education policy 2001. D. Ravitch, (Ed.) Washington, DC: Brookings Institution Press.

Texas Education Agency. (1996 April). The development of accountability systems nationwide and in Texas. Statewide Texas Educational Progress Study, Report No. 1. Austin: Texas Education Agency.

BEST COPY AVAILABLE

The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in Health Professions

Summary of the Symposium on Diversity in Health Professions
http://www.nap.edu/openbook.php?isbn=0309075645&page=280, copyright 2001, 2001 The National Academy of Sciences, all rights reserved

Treisman, U., & Fuller, E. (2001). Comment by Philip Uri Treisman and Edward J. Fuller, 2001. Brookings Papers on Education Policy 2001. D. Ravitch (Ed.). Washington, DC: Brookings Institution Press.

Tyack, D. (1974). The one best system: A history of American urban education. Cambridge, MA: Harvard University Press.

Valentine, J. (1987). The College Board and the school curriculum: A history of the College Board's influence on the substance and standards of American Education, 1900–1980. New York: College Entrance Examination Board.

EST COPY AVAILABLE